AD-A049 498 BROWN UNIV PROVIDENCE R I DIV OF ENGINEERING F/G 20/1

SIGNAL ESTIMATION AND SYSTEM IDENTIFICATION WITH LIMITED TIME D--ETC(U)

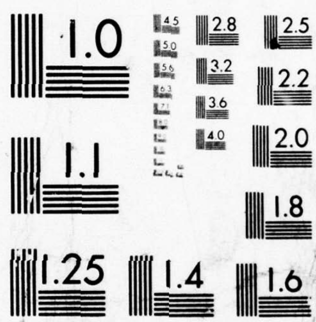
DEC 77 A E PEARSON AFOSR-75-2793

UNCLASSIFIED AFOSR-TR-78-0006 NL

| OF |

AD
A049498

END
DATE
FILMED
3-78
DDC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

AD A049498

DDC FILE COPY

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFOSR-TR-78-0006	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) SIGNAL ESTIMATION AND SYSTEM IDENTIFICATION WITH LIMITED TIME DATA		5. TYPE OF REPORT & PERIOD COVERED Interim rept.
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) A. E. Pearson		8. CONTRACT OR GRANT NUMBER(s) AFOSR-75-2793, NSF-ENG76-15000
9. PERFORMING ORGANIZATION NAME AND ADDRESS Brown University Division of Engeering Providence, Rhode Island 02912		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F 2304/A1 A1
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research/NM Bolling AFB DC 20332		12. REPORT DATE Dec 77 30p.
		13. NUMBER OF PAGES 27
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

DDC RECEIVED JAN 31 1978 F

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

With disturbances modeled by arbitrary solutions to a linear homogeneous differential equation, a deterministic theory is developed for parameter estimation using data over a limited time interval which has applications to signal estimation and system identification. Linear and certain classes of over

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

065 310

20. ABSTRACT (Continued)

nonlinear and time varying systems can be treated for identification purposes. The approach circumvents the need to estimate unknown initial conditions through the use of a certain projection operator.

AFOSR-TR- 78-0006

Signal Estimation and System Identification With Limited Time Data *

by

A. E. Pearson
Division of Engineering and
Lefschetz Center for Dynamical Systems
Brown University
Providence, Rhode Island 02912
U.S.A.

Summary

With disturbances modeled by arbitrary solutions to a linear homogeneous differential equation, a deterministic theory is developed for parameter estimation using data over a limited time interval which has applications to signal estimation and system identification. Linear and certain classes of nonlinear and time varying systems can be treated for identification purposes. The approach circumvents the need to estimate unknown initial conditions through the use of a certain projection operator.

* This research was supported in part by the Air Force Office of Scientific Research under grant AFOSR-75-2793B, and in part by the National Science Foundation under grant ENG76-15000.

I. INTRODUCTION

The success in using stochastic disturbance and measurement noise models with underlying Markov process representations is well documented for signal estimation and system identification [1]. Yet this does not preclude the possibility that alternative, i.e., nonstochastic, models might be advantageous in certain situations. This paper is concerned with one such potentially useful approach in which the disturbances are modeled as arbitrary solutions to a linear homogeneous differential equation of pre-selected order (r) over a fixed finite time interval, $0 \leq t \leq t_1$, and "one shot" signal estimation or system identification is to be undertaken based on the observed data on $[0,t_1]$. No assumptions are made regarding either the coefficients or the initial conditions for this differential equation, so that a variety of disturbances on $[0,t_1]$ can be represented by a moderately low order model, e.g., $r = 3$ or 4, if the observation time interval is of rather limited duration.

Although such disturbance models have been utilized in the past in connection with compensator design for the servomechanism problem, [2] and [3], or more recently by Davison [4] for a "compensator identification" problem with asymptotic tracking properties, the use of this model is quite different here. In the first place, the time interval $[0,t_1]$ is finite and, at least theoretically, may be arbitrarily short. Secondly, no restrictions are placed on the disturbance modes, i.e., they may be stable, unstable, or a mixture of both. Finally, no attempt is made to identify the initial conditions, either in the system model or in the disturbance model. This latter property results from the application of an annihilating filter (introduced in [5]), which zeroes the initial condition response of a linear system on a fixed time interval $[0,t_1]$. Relative to the results in [5], the present paper extends the class of models to include certain nonlinear and time-varying systems, allows for models in which the parameters enter nonlinearly, and includes a signal estimation problem as an application of the basic formulation. In addition, the parameters for the disturbance model are represented explicitly in this paper, rather than implicitly as in the case of [5].

The theoretical development of the approach is given in Section II starting from a basic differential operator model. Two formulations are presented depending on whether the model parameters are "separable" or not. In each case the parameter estimation problem is transformed into a certain kind of least squares fit with the pertinent functional obtained as an inner product over a subspace of the function space to which the observed data is assumed to belong. This subspace results from the use of the annihilating filter and serves to obviate the need to estimate the unknown initial conditions. It is then shown in Sections III and IV how particular identification problems can be translated into the basic formulations developed in Section II. Computational considerations are briefly discussed in Section V, but actual numerical results are reported elsewhere.

II. THEORETICAL DEVELOPMENT

It is shown in Sections III and IV how particular versions of the signal estimation and system identification problems can be viewed as finding a parameter vector $\theta = (\theta_1 .. \theta_p)$ which satisfies a differential operator equation of the generic form

$$P(D)v(t) - Q(D)g(t,\theta) = 0 \quad , \quad 0 \leq t \leq t_1 \tag{1}$$

where P and Q are polynomial matrices in the differential operator

$$D = \frac{d}{dt}$$

given by

$$P(D) = \sum_{i=o}^{n} P_i D^{n-i} \quad , \quad Q(D) = \sum_{i=o}^{n} Q_i D^{n-i}$$

and $(v(t), g(t,\theta))$ are column vector functions of the given data on $[0,t_1]$ and parameter vector θ as indicated. With respect to smoothness, $v(t)$ and $g(t,\theta)$ are assumed to be piecewise continuous functions of t on $[0,t_1]$ and are presumed not to depend on derivatives of the data; $g(t,\theta)$ is assumed to be continuously differentiable with respect to θ for each fixed t.

Definition The basic model (1) is said to be separable in the parameters if $g(t,\theta)$ admits to the representation

$$g(t,\theta) = V(t)f(\theta) \tag{2}$$

where $V(t)$ is a matrix valued function of the data and $f(\cdot)$ is a continuously differentiable vector valued function of the parameters with the single valued property:

$$f(\theta) = f(\theta^*) \text{ if and only if } \theta = \theta^* \tag{3}$$

for all θ and θ^*.

It will be shown that the computational burden is significantly less for separable models, and that the range of $f(\cdot)$ is generally of higher dimension than its domain.

Although system identification will be discussed more fully in Section IV, the following simple example will illustrate the above notation before continuing with the development.

Example 1 Consider the identification of parameters in the Mathieu equation with disturbance input $d(t)$:

$$\ddot{y}(t) + [\alpha_1 - \alpha_2 \cos \alpha_3 t]y(t) = \beta[u(t) + d(t)] \quad , \quad 0 \leq t \leq t_1 \qquad (4)$$

based on the observed input-output pair $[u(t),y(t)]$ on the observation interval $[0,t_1]$. Assume the disturbance model

$$\dot{d}(t) + \omega d(t) = (D + \omega)d(t) = 0 \quad , \quad 0 \leq t \leq t_1 \qquad (5)$$

where ω is a parameter which is to be identified along with the other system parameters. The disturbance $d(t)$ can be eliminated from (4) by operating on both sides with $(D + \omega)$. After rearranging terms the resulting differential equation can be expressed as

$$D^3 y(t) + [D^2 \; D \; 1] \begin{bmatrix} \omega y(t) \\ [\alpha_1 - \alpha_2 \cos\alpha_3 t]y(t) - \beta u(t) \\ \omega\{[\alpha_1 - \alpha_2 \cos\alpha_3 t]y(t) - \beta u(t)\} \end{bmatrix} = 0 \qquad (6)$$

which is of the form (1) with $P(D) = D^3$, $Q(D) = -\mathrm{Row}(D^2,D,1)$ and $\theta = (\alpha_1, \alpha_2, \alpha_3, \beta, \omega)$.

The model (6) is nonseparable in the parameters; however, it reduces to the separable case if the parameter α_3 is a known constant since the $g(t,\theta)$ vector can then be written as follows:

$$\begin{bmatrix} \omega y(t) \\ [\alpha_1-\alpha_2\cos\alpha_3 t]y(t) - \beta u(t) \\ \omega\{[\alpha_1-\alpha_2\cos\alpha_3 t]y(t) - \beta u(t)\} \end{bmatrix} = V(t)f(\tilde{\theta})$$

$$= \begin{bmatrix} y(t) & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & y(t) & -y(t)\cos\alpha_3 t & -u(t) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & y(t) & -y(t)\cos\alpha_3 t & -u(t) \end{bmatrix} \begin{bmatrix} \omega \\ \alpha_1 \\ \alpha_2 \\ \beta \\ \omega\alpha_1 \\ \omega\alpha_2 \\ \omega\beta \end{bmatrix} \qquad (7)$$

where $V(t)$ depends on the given input-output data, the parameter vector is now $\tilde{\theta} = (\alpha_1, \alpha_2, \beta, \omega)$ and $f(\tilde{\theta})$ satisfies the single-valued condition (3).

Given the polynomial matrix pair $[P(D),Q(D)]$ in the basic model (1), let a square polynomial matrix $F(D)$ be selected in the form

$$F(D) = \sum_{i=o}^{m} F_i D^{m-i} \quad , \; m \geq n \qquad (8)$$

with the integer m and coefficient matrices F_i chosen so that $F^{-1}(D)$ exists and $F^{-1}(D)[P(D),Q(D)]$ is a causal, i.e., proper, transfer function matrix. Then define an auxiliary error function $z(t) = z(t,\theta)$ implicitly through the solution to the differential operator equation

$$F(D)z(t) = P(D)v(t) - Q(D)g(t,\theta) \quad , \quad 0 \leq t \leq t_1 \quad . \qquad (9)$$

If (A,C) is an observable matrix pair for the homogeneous part of (9)†, then the solution for any particular initial condition x_o can be written as

$$z(t,\theta) = Ce^{At}x_o + h(t) - \mu(t,\theta) \quad , \quad 0 \leq t \leq t_1 \tag{10}$$

where $h(t)$ and $\mu(t,\theta)$ are the <u>zero state solutions</u> to

$$F(D)h(t) = P(D)v(t) \tag{11-a}$$

and

$$F(D)\mu(t,\theta) = Q(D)g(t,\theta) \tag{11-b}$$

respectively. In the case where the basic model (1) is separable in the parameters, the vector $\mu(t,\theta)$ can be written as

$$\mu(t,\theta) = M(t)f(\theta) \tag{12}$$

where $M(t)$ is the <u>zero state solution</u> to the matrix differential equation

$$F(D)M(t) = Q(D)V(t) \quad . \tag{13}$$

Given the polynomial matrix $F(D)$ and a corresponding observable matrix pair (A,C) for the homogeneous part of the auxiliary error equation (9), the annihilation filter relative to a fixed observation time interval $[0,t_1]$ is defined by the (noncausal) kernel function matrix

$$\mathscr{A}: \quad H(t,\tau) = I\delta(t-\tau) - Ce^{At}W^{-1}e^{A'\tau}C' \tag{14-a}$$
$$0 \leq t \leq t_1 \quad , \quad 0 \leq \tau \leq t_1$$

where $\delta(t)$ is the Dirac delta function, I is the identity matrix, W^{-1} is

---

† That is, all solutions to $F(D)z(t) = 0$ can be expressed by $z(t) = Cx(t) = Ce^{At}x_o$, $\dot{x}(t) = Ax(t)$, $x(o) = x_o \in \mathscr{R}^{\bar{n}}$, for an appropriate observable matrix pair (A,C) with minimal dimension state space $(\bar{n})$, and where e^{At} is the state transition matrix for A.

inverse of the observability Gramian for the pair (A,C) defined by

$$W = \int_0^{t_1} e^{A't} C' C e^{At} dt \quad , \tag{14-b}$$

and prime denotes transpose. If $\mathcal{Z}$ denotes the Hilbert space of all vector valued square integrable functions $z(t)$ which are possible solutions to (9) on $[0,t_1]$, and $\mathcal{Z}_o$ the linear subspace spanned by the columns of Ce^{At},† then it is easy to show that $\mathcal{A}$ is a projection operator on $\mathcal{Z}$ with range contained in $(\mathcal{Z}-\mathcal{Z}_o)$ which possesses the annihilation property

$$\mathcal{A}(Ce^{At}x_o) = 0 \quad , \quad 0 \leq t \leq t_1 \quad , \text{ for all } \; x_o \varepsilon \mathcal{R}^{\bar{n}} \quad . \tag{15}$$

This property follows immediately from the definition of $\mathcal{A}$ in (14). The fact that $\mathcal{A}$ is a projection follows upon noting that $\mathcal{A}(\mathcal{A}) = \mathcal{A}$.

Operating on both sides of (10) with $\mathcal{A}$ yields

$$\tilde{z}(t,\theta) = \tilde{h}(t) - \tilde{\mu}(t,\theta) \; , \; 0 \leq t \leq t_1 \tag{16}$$

where

$$\tilde{h}(t) = h(t) - Ce^{At}W^{-1}\int_0^{t_1} e^{A'\tau} C' h(\tau) d\tau \tag{17-a}$$

and

$$\tilde{\mu}(t,\theta) = \mu(t,\theta) - Ce^{At}W^{-1}\int_0^{t_1} e^{A'\tau} C'\mu(\tau,\theta) d\tau \tag{17-b}$$

† Thus, $\mathcal{Z}_o = \{z(t) = Ce^{At}x_o \; , \; 0 \leq t \leq t_1 : x_o \varepsilon \mathcal{R}^{\bar{n}}\}$

are the projections of $h(\cdot)$ and $\mu(\cdot,\theta)$ down into the subspace $(\mathcal{Z}-\mathcal{Z}_0)$. Defining the inner product functional $J_1(\theta)$ by

$$J_1(\theta) = \langle\tilde{z}(\theta), \tilde{z}(\theta)\rangle = \int_0^{t_1} \tilde{z}'(t,\theta)\tilde{z}(t,\theta)dt \tag{18}$$

it follows from the above development that any value of θ which satisfies the basic differential operator model (1) is also a solution to the nonlinear transcendental equation

$$J_1(\theta) = 0 \quad . \tag{19}$$

Conversely, any value of θ which satisfies (19) is a candidate for a value of the parameter vector θ satisfying (1).

A straightforward calculation of the quantities involved† shows that $J_1(\theta)$ can be represented by

$$\begin{aligned} J_1(\theta) = \int_0^{t_1} h'(t)h(t)dt - \eta' W^{-1}\eta - 2\int_0^{t_1} h'(t)\mu(t,\theta)dt \\ + 2\eta' W^{-1}\nu(\theta) + \int_0^{t_1} \mu'(t,\theta)\mu(t,\theta)dt - \nu'(\theta)\, W^{-1}\nu(\theta) \end{aligned} \tag{20}$$

where $h(t)$ and $\mu(t,\theta)$ are given by (11) and the vectors $(\eta, \nu(\theta))$ are defined by

$$\eta = \int_0^{t_1} e^{A't}C'h(t)dt \tag{21-a}$$

$$\nu(\theta) = \int_0^{t_1} e^{A't}C'\mu(t,\theta)dt \quad . \tag{21-b}$$

† Specifically, substituting (16) and (17) into (18), and simplifying the resulting expression.

An iterative solution to (19) can then be attempted via any suitable parameter search method, with the gradient $\nabla J_1(\theta)$ and any higher order derivatives of $J_1(\theta)$ computable by causal relations involving the input-output data on $[0,t_1]$, relative to any particular value of θ . Equation (11-b) must be integrated anew (from the zero state) for each value of θ in the course of constructing a sequence $\{\theta(n)\}$ as a possible solution to (19). More generally, the limit of a minimizing sequence for $J_1(\theta)$ represents a kind of least squares estimate of the parameter vector θ which may, or may not, be unique depending on the nature of the model (1) and the observed data on $[0,t_1]$.

In the case of a model which is separable in the parameters, $\mu(t,\theta)$ is given by (12) and $\nu(\theta)$ in (21-b) can be written as

$$\nu(\theta) = Nf(\theta) \tag{22}$$

where the matrix N is defined by

$$N = \int_0^{t_1} e^{A't}C'M(t)dt \quad . \tag{23}$$

In this case, the function $J_1(\theta)$ reduces to the following explicitly defined function of θ :

$$J_2(\theta) = \alpha - 2\zeta'f(\theta) + f'(\theta)\Phi f(\theta) \tag{24}$$

where (α,ζ,Φ) are defined by

$$\alpha = \int_0^{t_1} h'(t)h(t)dt - \eta'W^{-1}\eta \tag{25-a}$$

$$\zeta = \int_0^{t_1} M'(t)h(t)dt - N'W^{-1}\eta \tag{25-b}$$

$$\Phi = \int_0^{t_1} M'(t)M(t)dt - N'W^{-1}N \quad .^{\dagger} \tag{25-c}$$

---

†The matrix Φ is actually the Gram matrix for the column vector functions of $\tilde{M}(t) = \mathscr{A}(M)$; as such, Φ is necessarily symmetric and non-negative definite regardless of the data on $[0,t_1]$.

Although $J_1(\theta)$ and $J_2(\theta)$ are equivalent positive definite functionals of θ whose zero values reflect possible values for the unknown parameter vector θ satisfying the basic model (1), the computational advantages of (24) for the separable-in-the-parameters case should be evident inasmuch as $J_2(\theta)$ is an explicitly defined function of θ, while $J_1(\theta)$ is defined only implicitly. This means that once the differential equations (11-a) and (13) are integrated (from the zero state) and the quantities $(\eta, N, \alpha, \zeta, \Phi)$ computed from (21-a), (23) and (25), there are no further integrations needed involving the data collected on $[0,t_1]$. Moreover, sufficient conditions for the uniqueness of solutions to the parameter estimation problem can be stated more specifically as follows.

Assertion In the case of the basic model (1) and (2) which is separable in the parameters, a minimizing value θ^* for the positive definite function $J_2(\theta)$ in (24) is a least squares estimate of the parameter vector θ which is unique if (as a sufficient condition) the given data is such that the columns of $Q(D)V(t)$ are linearly independent functions on $0 \leq t \leq t_1$.

The fact that a minimizing value θ^* for $J_2(\theta)$ is a least squares estimate follows from the consideration that if $J_2(\lambda)$, $\lambda = f(\theta)$ is minimized over λ, rather than θ, the necessary condition, $\nabla J_2(\lambda) = 0$, is seen to be

$$\frac{1}{2} \nabla J_2(\lambda) = \zeta - \Phi\lambda = 0$$

which in turn can be seen to be the normal equations for the projected auxiliary error function $\tilde{z}(t,\theta) = \tilde{h}(t) - \tilde{M}(t)f(\theta) = \tilde{h}(t) - \tilde{M}(t)\lambda$ in (16). Since $\Phi = \int_0^{t_1} \tilde{M}'(t)\tilde{M}(t)dt$ is the Gram matrix for the column vector functions of $\tilde{M}(\cdot)$, a unique solution to these normal equations is obtained if and only if the columns of $\tilde{M}(t)$ are linearly independent on $[0,t_1]$. Now the columns of $\tilde{M}(t)$

are wholly contained in the subspace $(\mathcal{J} - \mathcal{J}_0)$ and can be represented by $(F^{-1}(D)Q(D)V(t)\lambda)^{\perp}$ where the symbol $(\ )^{\perp}$ denotes the orthogonal complement, or projection, of the function $F^{-1}(D)Q(D)V(t)\lambda$ in the subspace $(\mathcal{J} - \mathcal{J}_0)$. Since $\mathcal{J}_0$ is the null space for $F(D)$, it follows that linear dependence, or independence, of the columns of $\tilde{M}(\cdot)$ cannot be destroyed, or altered, by operating on $(F^{-1}(D)Q(D)V(t)\lambda)^{\perp}$ with $F(D)$. By this argument, the sufficiency proof for uniqueness in the Assertion is established. This condition is also necessary when the function $f(\theta)$ is just θ since $J_2(\theta)$ is then a positive definite quadratic form in θ. This will be the case for the signal estimation problem in Section III, as well as any problem for which the unknown parameters enter linearly in the basic model (1) - (2).

III. Signal Estimation

Let an observed scalar signal $y(t)$ on $[0,t_1]$ be represented by

$$y(t) = s(t) + d(t) \tag{26}$$

where the useful signal $s(t)$ and the disturbance $d(t)$ are assumed to be modeled by the differential operator equations

$$s(t): \qquad A(D)s(t) = B(D)u(t) \tag{27-a}$$

$$d(t): \qquad T(D,\theta)d(t) = \sum_{i=o}^{r} \theta_i D^{r-i} d(t) = 0. \tag{27-b}$$

$$\theta_o = 1$$

In the model (27-a) for the signal $s(t)$ which is to be estimated, the differential operator polynominals $(A(D),B(D))$ and the deterministic signal $u(t)$, $0 \leq t \leq t_1$, are assumed to be given, but the initial conditions are unknown. The order (r) of the disturbance model (27-b) is assumed to be specified, but the coefficients $\theta = (\theta_1 \cdot\cdot \theta_r)$ and initial conditions are unknown.

Operating on both sides of (26) with $T(D,\theta)A(D)$, noting (27), and rearranging terms leads to the differential operator equation

$$0 = D^r[B(D)u(t) - A(D)y(t)] - [D^{r-1} \cdots D \; 1][A(D)y(t) - B(D)u(t)] \begin{bmatrix} \theta_1 \\ \cdot \\ \cdot \\ \cdot \\ \theta_r \end{bmatrix} \tag{28}$$

which is of the form (1) and (2), i.e., separable in the parameters, with the parameter vector $\theta = \mathrm{Col}(\theta_1 \cdots \theta_r)$ entering linearly. A least squares estimate of θ is obtained either by minimizing the quadratic functional $J_2(\theta)$ with $f(\theta) = \theta$ in (24), or by solving the "normal" equations

$$\zeta = \Phi\theta \tag{29}$$

directly, where ζ and Φ are given in (25) after first integrating the differential equations (11-a) and (13) for this problem, i.e., obtaining the zero state solutions to

$$F(D)h(t) = D^r B(D)u(t) - D^r A(D)y(t) \tag{30-a}$$

and

$$F(D)M(t) = [D^{r-1} \cdots D\ 1][A(D)y(t) - B(D)u(t)]. \tag{30-b}$$

Here $h(t)$ is a scalar while $M(t)$ is a row vector $(M_1(t) \cdot\cdot M_r(t))$.

Let $\hat{\theta}$ denote the resulting least squares estimate and define an annihilation filter $\mathscr{A}_\theta$ by its impulse response function $H_\theta(t,\tau)$ analogous to (14):

$$H_\theta(t,\tau) = \delta(t-\tau) - C_\theta e^{A_\theta t} W_\theta^{-1} e^{A_\theta' \tau} C_\theta', \quad 0 \leq t,\tau \leq t_1 \tag{31}$$

where (A_θ, C_θ) is an observable pair for the disturbance model (27-b) with $\theta = \hat{\theta}$, e.g.

$$A_\theta = \begin{bmatrix} 0 & 1 & & 0 \\ \cdot & & \ddots & \cdot \\ \cdot & & & \cdot \\ 0 & 0 & \cdots & 1 \\ -\hat{\theta}_r & \cdot & \cdot\ \cdot & -\hat{\theta}_1 \end{bmatrix}, \qquad C_\theta = \mathrm{Row}(1,0 \cdot\cdot 0). \tag{32}$$

The signal estimate is then defined by

$$\begin{aligned} \hat{s}(t) &= \mathscr{A}_\theta(y(t)) \qquad\qquad 0 \leq t \leq t_1 \\ &= y(t) - C_\theta e^{A_\theta t} W_\theta^{-1} \int_0^{t_1} e^{A_\theta' \tau} C_\theta' y(\tau)d\tau \end{aligned} \tag{33}$$

which is the projection of the observed data $y(t)$ down into the subspace obtained by deleting all linear combinations of the disturbance modes identified via the solution to the normal equations (29). Using the function space norm, the error in this estimate can be found to have the following representation:

$$||s - \hat{s}||^2 = d_o' W_\theta d_o$$

where

$$d_o = W_\theta^{-1} \int_o^{t_1} e^{A_\theta' t} C_\theta' s(t)dt.$$

It is clear that the estimate $\hat{s}(t)$ is precisely the signal $s(t)$ under the conditions: (i) the models in (27) are correct, (ii) det $\Phi \neq 0$, and (iii) $s(\cdot)$ is orthogonal to each of the basis functions comprising $d(\cdot)$, i.e. $d_o = 0$.

Example 2: Fourier Series Model for s(t)

Given the time interval $[0,t_1]$, define a fundamental frequency ω_o by

$$\omega_o = \frac{2\pi}{t_1}$$

and select the polynomials (A(D),B(D)) in (27-a) according to

$$A(D) = D(D^2+\omega_o^2)(D^2+4\omega_o^2)\cdot\cdot(D^2+m^2\omega_o^2)$$
$$B(D) = 0.$$

Then the solution to the signal model (27-a) is the truncated Fourier series

$$s(t) = \sum_{n=-m}^{m} S_n e^{jn\omega_o t} \quad , \quad 0 \leq t \leq t_1.$$

IV. System Identification

A number of examples will be given in this section to illustrate the basic theory for system parameter identification.

Example 3: Linear Differential Systems

Let the input-output relation for a class of linear systems be characterized in the absence of input or measurement disturbances by the differential operator equation

$$A(D,\omega)y(t) = B(D,\omega)u(t) \tag{34}$$

where $$A(D,\omega) = \sum_{i=o}^{n} A_i(\omega)D^{n-i} \quad , \quad B(D,\omega) = \sum_{i=o}^{n} B_i(\omega)D^{n-i}.$$

The coefficient matrices $(A_i(\omega),B_i(\omega)), 0 \leq i \leq n$, are assumed to be given functions of a parameter vector $\omega = (\omega_1 \cdot\cdot \omega_p)$. Define a vector valued function $f(\omega)$ with components $f_i(\omega)$ selected so as to reflect all the various distinct ways in which the parameters enter into the $A_i(\omega)$ and $B_i(\omega)$, *i.e. linear, multiplicatively*, etc. It is then easy to see how to define quantities $\{P(D),Q(D), v(t),V(t)\}$ with $v(t)$ and $V(t)$ depending on the input-output pair $[u(t),y(t)]$ such that the following decomposition holds:

$$\begin{aligned} A(D,\omega)y(t) - B(D,\omega)u(t) &= P(D)v(t) - Q(D)V(t)f(\omega) \\ &= 0. \end{aligned} \tag{35}$$

Now suppose the input and output are corrupted by additive disturbances such that the observed output $y(t)$ includes a measurement disturbance $d_1(t)$ according to

$$y(t) = y_o(t) + d_1(t)$$

and the actual input to the system is the signal

$$u_o(t) = u(t) + d_2(t)$$

while only u(t) can be directly observed. Including these disturbances in (34) implies:

$$A(D,\omega)[y(t) - d_1(t)] = B(D,\omega)[u(t) + d_2(t)]. \quad (36)$$

If $d_1(t)$ and $d_2(t)$ are assumed to be arbitrary solutions to differential equations of the form (27-b) on the observation interval $[0,t_1]$, it makes no difference whether all such disturbance modes are present in the input, that they corrupt only the output, or some combination thereof, since operating on both sides of (36) with $T(D,\theta)$ yields

$$T(D,\theta)A(D,\omega)y(t) = T(D,\theta)B(D,\omega)u(t). \quad (37)$$

With the decomposition (35) already defined for the system without disturbances, the analogous decomposition for the model (37) with disturbances modeled by (27-b) can be easily shown to be as follows:

$$0 = D^r P(D)v(t)$$

$$-[D^{r-1}P(D)v(t)\cdots P(D)v(t) \mid D^r Q(D)V(t) \mid D^{r-1}Q(D)V(t)\cdots Q(D)V(t)] \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_r \\ \hline f(\omega) \\ \hline \theta_1 f(\omega) \\ \vdots \\ \theta_r f(\omega) \end{bmatrix} \quad (38)$$

This differential operator equation is in the form of the separable-in-the-parameters model (1) and (2) with parameter vector $\tilde{\theta} = (\theta_1\cdots\theta_r,\omega_1\cdots\omega_p)$. The vector function $\tilde{f}(\tilde{\theta})$ on the right side of (38) satisfies the single-valued property (3) if the original function $f(\omega)$ in (35) satisfies this property. This will normally be the case for a properly parametrized model.

It is apparent from (37) that the disturbances can be equivalently interpreted as uncontrollable modes. In contrast with the formulation in [5], the parameters $(\theta_1\cdots\theta_r)$ for these modes are modeled explicitly here, rather than implicitly. This avoids the polynomial factorization discussed in [5] when the disturbances are modeled implicitly.

As specific illustration of this notation consider the following state equations for a helicopter in longitudinal motion as given by Narendra and Tripathi [6]:

$$\dot{x}(t) = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & \omega_2 & a_{33} & \omega_3 \\ 0 & 0 & 1 & 0 \end{bmatrix} x(t) + \begin{bmatrix} b_{11} & b_{12} \\ \omega_1 & b_{22} \\ b_{31} & b_{32} \\ 0 & 0 \end{bmatrix} u(t)$$

$$y(t) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} x(t) .$$

According to Narendra and Tripathi the (b_{21}, a_{32}, a_{34}) entries (designated by $(\omega_1, \omega_2, \omega_3)$ above) vary significantly over the airspeed range 60-170 knots, while the remaining entries maintain relatively constant values. In addition, the vertical velocity (x_2) is difficult to measure, and the pitch rate (x_3) might entail high frequency measurement noise. Hence, horizontal velocity (x_1) and the pitch angle (x_4) are regarded here as the measurable outputs, together with the measurable collective (u_1) ahd longitudinal cyclic (u_2) pitch control input variables, and $\omega = (b_{21}, a_{32}, a_{34})$ is regarded as the parameter vector for identification. Eliminating the states (x_2, x_3) in order to obtain an input-output relation of the form (34), and rearranging this equation into the requisite form (35), the quantities (P,Q,v,V,f) are found to be as follows:

$$P(D) = P_o D^3 + P_1 D^2 + P_2 D + P_3 , \qquad P_o = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$Q(D) = \begin{bmatrix} D & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad v(t) = \begin{bmatrix} -y(t) \\ u(t) \end{bmatrix} \qquad f(\omega) = \begin{bmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \\ \omega_1 \omega_2 \end{bmatrix}$$

$$V(t) = \begin{bmatrix} 0 & -a_{23}y_2(t) & -y_2(t) & 0 \\ 0 & y_3(t) & a_{22}y_2(t) & -u_1(t) \\ -a_{12}u_1(t) & 0 & 0 & 0 \end{bmatrix}$$

$$y_3(t) \triangleq -a_{21}y_1(t) - a_{24}y_2(t) - b_{22}u_2(t)$$

The (P_1,P_2,P_3) matrices depend on the fixed (a_{ij},b_{ij}) values.

Example 4: Time Lag Systems

Let the model (34) in Example 3 be modified to include an unknown pure time delay α:

$$A(D,\omega)y(t) = B(D,\omega)u(t-\alpha), \quad \alpha \leq \alpha_{max}.^{\dagger} \tag{39}$$

It is clear that a decomposition such as (35) will not exist in this instance so that the time-delay system is not separable in the delay parameter. However, input-output disturbances similar to those included in (36) can still be incorporated leading to (37) with u(t) replaced by $u(t-\alpha)$. The basic model (1) will then apply in this case for appropriately defined quantities $\{P,Q,v,g\}$ with the parameter vector $\tilde{\theta}$ comprised of (θ,ω,α).

As a specific illustration of the notation involved, consider a single input-single output system with

$$A(D,\omega) = D^2+\omega_1D+\omega_2 \quad , \quad B(D,\omega) = \omega_3$$

$$T(D,\theta) = D+\theta \ .$$

The input-output relation (39) with disturbances is then

$$(D^2+\omega_1D+\omega_2)[y(t)-d_1(t)] = \omega_3[u(t-\alpha)+d_2(t-\alpha)].$$

Using a first order disturbance model for both $d_1(t)$ and $d_2(t)$, the above is equivalent to the following differential delay operator equation (obtained after some rearrangement):

---

† Although α is unknown, its value is assumed to be bounded by a given number α_{max} as indicated.

$$D^3y(t) + [D^2 \; D \; 1]\begin{bmatrix}(\omega_1+\theta)y(t) \\ (\omega_2+\omega_1\theta)y(t) - \omega_3 u(t-\alpha) \\ \omega_2\theta y(t) - \omega_3\theta u(t-\alpha)\end{bmatrix} = 0 \tag{40}$$

which is of the form (1) with parameter vector $\tilde{\theta} = (\theta,\omega_1,\omega_2,\omega_3,\alpha)$.

Even though the "state" of the models (39) and (40) is infinite dimensional, the basic theory of Section II leading to the functional $J_1(\theta)$ in (20) is still valid. However, the given data must be assumed to include the past input $u(t)$ for $t\varepsilon[-\alpha_{max},0]$ in addition to the input-output pair $[u(t),y(t)]$ on $[0,t_1]$. This will ensure the existence of a solution to (19) for computational considerations.

Example 5: Hammerstein Model

Consider the scalar system with observed input-output pair $[u(t),y(t)]$ on $[0,t_1]$:

$$A(D,\omega)[y(t) + d_1(t)] = B(D,\omega)[\sum_{i=1}^{k} \alpha_i (u(t))^i + d_2(t)] \tag{41}$$

where $d_1(t)$ and $d_2(t)$ are unknown disturbances. This is a Hammerstein model [7] with output disturbance $d_1(t)$ and intermediate input disturbance $d_2(t)$ entering the system after the zero memory nonlinearity characterized by the parameters $\alpha = (\alpha_1 \cdot\cdot \alpha_k)$. Assuming the disturbance model (27-b) for $d_1(t)$ and $d_2(t)$, the model (41) is modified to

$$T(D,\theta)A(D,\omega)y(t) = T(D,\theta)B(D,\omega)\sum_{i=1}^{k}\alpha_i(u(t))^i. \tag{42}$$

It is evident that with appropriate definitions of the quantities (P,Q,v,V,f), (42) can be arranged in the form of the separable-in-parameters model (1) and (2) with parameter vector $\tilde{\theta} = (\theta,\omega,\alpha)$, given any particular $(A(D,\omega),B(D,\omega))$ polynomials and integer r in (27-b).

Example 6: Peak Output System

Consider a scalar system characterized by the model

$$A(D,\omega)[c(t)+d_1(t)] = B(D,\omega)[u(t)+d_2(t)] \tag{43-a}$$

$$y(t) = \alpha - (c(t))^2 \quad ^{\dagger} \tag{43-b}$$

where $d_1(t)$ and $d_2(t)$ are again disturbances of the type modeled in (27-b), $c(t)$ is an intermediate output signal preceding a parabolic nonlinearity with unknown peak value α, and $[u(t),y(t)]$ is the observed input-output pair on $[0,t_1]$. Incorporating the disturbance model (27-b) into (43-a) and replacing $c(t)$ by $\pm[\alpha-y(t)]^{1/2}$:

$$\pm T(D,\theta)A(D,\omega)[\alpha-y(t)]^{1/2} = T(D,\theta)B(D,\omega)u(t). \tag{44}$$

Given polynomials $(A(D,\omega),B(D,\omega))$ and an integer r for the disturbance model (27-b), it is clear that (44) can be arranged into the form of the nonseparable basic model (1) with parameter vector $\tilde{\theta} = (\theta,\omega,\alpha)$. For example, taking the second order system with first order disturbances:
$A(D,\omega) = D^2+\omega_1 D+\omega_2$, $B(D,\omega) = \omega_3$, $T(D,\theta) = D+\theta$, the rearrangement of (44) in this case laads to

$$0 = [D^3 \; D^2 \; D \; 1]\begin{bmatrix} \pm[\alpha-y(t)]^{1/2} \\ \pm(\theta+\omega_1)[\alpha-y(t)]^{1/2} \\ \pm(\theta\omega_1+\omega_2)[\alpha-y(t)]^{1/2} - \omega_3 u(t) \\ \pm\; \theta\omega_2[\alpha-y(t)]^{1/2} - \theta\omega_3 u(t) \end{bmatrix} \tag{45}$$

Here, the term $P(D)v(t)$ in (1) is absent so that the inner product terms involving $h(t)$ and η in (20) are zero, leaving

---

†Notice that a shift in the parabola according to $y = \alpha-(c-c^*)^2$ is unnecessary since the disturbance model for $d_1(t)$ includes c^* as a special case.

$$J_1(\tilde{\theta}) = \int_0^{t_1} [\mu(t,\tilde{\theta})]^2 dt - \nu'(\tilde{\theta})W^{-1}\nu(\tilde{\theta}) = 0 \qquad (46)$$

where $\tilde{\theta} = (\theta,\omega_1,\omega_2,\omega_3,\alpha)$. The ambiguity in sign for the term $\pm(\alpha-y)^{1/2}$ in (44) and (45) cannot be resolved and may cause nonuniqueness in finding the zeros of (46) even if the input is sufficiently active to excite all the modes in the system on $[0,t_1]$. In any case, the initial guess for the peak output should be chosen to satisfy

$$\alpha(1) > \max_{0 \leq t \leq t_1} y(t)$$

in the construction of any sequence $\{\tilde{\theta}(n)\}$, $n = 1,2\cdots$, directed at solving (46).

V. Computational Considerations

Assuming the basic model (1) has been obtained for any particular application, it is necessary to choose the polynomial matrix $F(D)$ in (8) of sufficiently high order so that (11-a) and (11-b), or (13), can be integrated without involving derivatives of the data on the right hand side of these equations. Apart from $\det F(D) \neq 0$, the selection of $F(D)$ is quite unrestricted and the modes of $F(D)$ can be selected as either stable or unstable since all computations are confined to the finite interval $[0, t_1]$. In this selection it is possible to take advantage of cancelling various modes in $P(D)$ and $Q(D)$, i.e. cancellation of poles and zeros in $F^{-1}(D)[P(D), Q(D)]$, in order to simplify the computations for $h(t)$, $\mu(t,\theta)$ or $M(t)$. However, it is necessary to include any such cancelled modes in the computation of Ce^{At} for the determination of W in (14b) and in the computation of ($\eta,\nu(\theta)$ or N) in (21) and (23). In this connection it should be evident that Ce^{At} and W^{-1} can be computed offline and stored for subsequent online computations as input-output data is presented.

Various simplifications in integrating the required differential equations can be gleened by a careful comparison of the equations involved. For example, comparing (30a) and (30b) which are needed to obtain the normal equations (29) in the signal estimation problem, it is evident that the components of $M(t)$ can be obtained by pure integrations of $h(t)$, viz.

$$M_1(t) = -\int_0^t h(\tau)d\tau \ , \ M_i(t) = \int_0^t M_{i-1}(\tau)d\tau \ , \ 0 \le t \le t_1 \qquad (47)$$
$$i = 2 \cdots r$$

Similar simplifications in computing the columns of $M(t)$ will be evident for the integrations arising from the basic model (38) in Example 3, as well as in the other examples.

Concerning the choice in the order (r) of the disturbance model (27-b), it could be argued that such a choice can never be made with any degree of certainty since disturbances are, by nature, unknown. Two comments are appropriate in this connection: (i) The **minimization** of either $J_1(\theta)$ in (20), or $J_2(\theta)$ in (24), represents a kind of generalized least squares estimate of the parameter vector θ appearing in the basic model (1) or (1)-(2), since (20) and (24) stem from the Hilbert space norm of the projected error function $\tilde{z}(t,\theta)$ in (16). (ii) If the minimal value of $J_1(\theta)$ or $J_2(\theta)$ exceeds a certain small value ε indicating inadequacy of the model, a natural course of action is to change the value of r and recompute the new functional. If $J(\theta;r)$ denotes the functional corresponding to a certain value of r in (27-b), a reasonable procedure to follow might be to start with a large value of r, say r_{max}, compute the necessary quantities for $r = r_{max}$, then note that the analogous quantities for $r < r_{max}$ can be obtained by pure integrations of the former in many cases. To illustrate this point, consider the quantities $h(t) = h(t;r)$ and $M(t) = M(t;r)$ in (30-a) and (30-b) for the signal estimation problem. Assuming $F(D)$ is chosen so that $h(t;r_{max})$ and $M(t;r_{max})$ are causally related to the given data, and assuming the same $F(D)$ for $r < r_{max}$, it is seen that

$$h(t;r) = D^{r-r_{max}}\, h(t;r_{max})$$

which is an $(r_{max} - r)$ fold integral of $h(t;r_{max})$. This consideration coupled with the observations in (47) can significantly reduce the number of integrations when considering various orders for the disturbances model (27-b).

As a final remark concerning the minimization of $J_2(\theta)$ in (24) for models which are separable in the parameters, a choice basically exists between minimizing J_2 over $\lambda = f(\theta)$, or minimizing directly over θ, in view of the single valued property (3). The advantage of the former is that J_2 is quadratic in

λ while it is generally highly nonlinear in θ. However, this advantage is offset by the fact that the dimension of λ may be very much greater than θ as can be seen from the various separable examples presented in the preceding sections.

VI. Concluding Remarks

The formulations of the signal estimation and system identification problems presented in this paper use models for the disturbances which are deterministic homogeneous linear differential equations of preselected order (r) on a finite time interval $[0,t_1]$. The shorter the time interval, the more realistic will be the assumption that the disturbances can be so modeled. On the other hand, the time interval must be long enough so that the data contains enough information to reflect the values of the parameters to be determined in the basic model. This can be put into somewhat more specific terms by referring to particular examples, viz. the matrix Φ in the normal equations (29) for the signal estimation problem must be sufficiently positive definite to yield a unique solution. A larger t_1 enhances this possibility since Φ is a Gram matrix for functions in the space $(\mathcal{Z} - \mathcal{Z}_0)$.

The examples for illustrating the basic model (1) are surely not exhaustive. The Fourier series model in Example 2 is a natural model for the signal estimation problem; yet there may exist applications for using the more general model of (27-a). The nonlinear examples for system identification referred to tandem interconnections of zero memory nonlinearities and linear dynamic subsystems. However, other identification problems can be handled such as the Duffing equation with input disturbance d(t):

$$\ddot{y}(t) + \omega_1 y(t)\dot{y}(t) + \omega_2 y(t) + \omega_3 y^3(t) = \omega_4[u(t) + d(t)].$$

In differential operator notation, this can be written as

$$D^2 y(t) + 1/2\ \omega_1 D(y^2(t)) + \omega_2 y(t) + \omega_3 y^3(t) = \omega_4[u(t)+d(t)]$$

having noted that $y\dot{y} = 1/2D(y^2)$. Employing the disturbance model (27-b) and proceeding as in the previously cited examples will lead to a model which is separable in the parameters. The Van der Pol equation can be similarly handled.

Finally, although most of these examples have involved time invariant systems, it should be clear from Example 1 that certain time varying systems can be placed in the form of (1).

No examples involving numerical calculations have been given in this paper. However, a number of computer simulations have been carried out for linear system identification, both fixed and time varying. Some of these are reported in Section V of [5], while others are contained in a recent thesis by Chin [8]. The results of the latter, which includes an aerospace adaptive control application with time varying parameters, will be reported elsewhere.

References

1. IEEE Trans. Auto. Control: Special Issue on System Identification and Time Series Analysis. AC-19, December (1974).

2. C. D. Johnson: Accomodation of External Disturbances in Linear Regulator and Servomechanism Problems. IEEE Trans. on Auto.Control, AC-16, 635-645 (1971).

3. E. J. Davison: The Output Control of Linear Time Invariant Multivariable Systems With Unmeasurable Arbitrary Disturbances. IEEE Trans. on Auto. Control, AC-17, 621 - 630 (1972).

4. E. J. Davison: Multivariable Tuning Regulators: The Feedforward and Robust Control of a General Servomechanism Problem. IEEE Trans. on Auto. Control, AC-21, 35 - 47 (1976).

5. A. E. Pearson: Finite Time Interval Linear System Identification Without Initial State Estimation. Automatica 12, 577 - 587 (1976).

6. K. S. Narendra and S. S. Tripathi: Identification and Optimization of Aircraft Dynamics. J. Aircraft, 10, 193-199 (1973).

7. P. G. Gallman: A Comparison of Two Hammerstein Model Identification Algorithms. IEEE Trans. on Auto. Control, AC-21, 124 - 126 (1976).

8. Y. K. Chin: Theoretical and Computational Considerations for Finite Time Interval System Identification Without Initial State Estimation. Ph.D. Thesis, Brown University (1977).